



A study on local photometric models and their application to robust tracking

Michèle Gouiffès, Christophe Collewet, Christine Fernandez-Maloigne, A. Trémeau

► To cite this version:

Michèle Gouiffès, Christophe Collewet, Christine Fernandez-Maloigne, A. Trémeau. A study on local photometric models and their application to robust tracking. *Computer Vision and Image Understanding*, 2012, 116 (8), pp.896-907. 10.1016/j.cviu.2012.04.002 . hal-00726513

HAL Id: hal-00726513

<https://inria.hal.science/hal-00726513>

Submitted on 30 Aug 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A study on local photometric models and their application to robust tracking.

M. Gouiffès⁽¹⁾, C. Collewet⁽²⁾, C. Fernandez⁽³⁾, A. Trémeau⁽⁴⁾

(1) michele.gouiffes@ief.u-psud.fr

Institut d'Electronique Fondamentale CNRS UMR 8622 Université

Paris-Sud 11, 91405 ORSAY cedex

(2) IRSTEA and INRIA, Rennes, France

(3) XLIM, Université de Poitiers, France

(4) Laboratoire Hubert Curien, St Etienne, France

Abstract

Since modeling reflections in image processing is a difficult task, most computer vision algorithms assume that objects are Lambertian and that no lighting change occurs. Some photometric models can partly answer this issue by assuming that the lighting changes are the same at each point of a small window of interest. Through a study based on specular reflection models, we explicit the assumptions on which these models are implicitly based and the situations in which they could fail.

This paper proposes two photometric models, which compensate for specular highlights and lighting variations. They assume that photometric changes vary smoothly on the window of interest. Contrary to classical models, the characteristics of the object surface and the lighting changes can vary in the area being observed. First, we study the validity of these models with respect to the acquisition setup: relative locations between the light source, the sensor and the object as well as the roughness of the surface. Then,

these models are used to improve feature points tracking by simultaneously estimating the photometric and geometric changes. The proposed methods are compared to well-known tracking methods robust to affine photometric changes. Experimental results on specular objects demonstrate the robustness of our approaches to specular highlights and lighting changes.

Keywords: Computer vision, robust feature point tracking, local photometric models.

1. Introduction

Computer vision has recently emerged in many fields such as mobile robotics [1], visual inspection, in surgical, industrial, agricultural, spatial or underwater domains [2], *i.e.* in various natural environments. For such practical applications, one of the crucial problems lies in the robustness of the low level algorithms with respect to some critical acquisition conditions: blurred images, acquisition noise, illumination changes, reflections. High level algorithms such as 3D reconstruction, active vision or visual servoing can be efficiently improved by increasing the robustness of the spatial and temporal matching process.

This paper addresses more precisely the problem of robust feature tracking with respect to lighting changes and specular highlights. The issue can be tackled by extracting salient features in the image, such as edges [3, 4], corners [5], lines [6]. It becomes far more complicated in most natural environments when only points are likely to be detectable. However, tracking those features is not trivial since it relies on the luminance of the neighbor pixels which are highly sensitive to photometric variations. The seminal

works in the domain are due to Lucas and Kanade [7, 8] (KLT) who assume the luminance constancy [9] in order to compute the translation motion of each point.

This work has been extended to more comprehensive motion models: affine [10], quadratic [11] and homographic [12, 13], and is still considered to be powerful [14]. The tracking of planar patterns can also be implemented by an efficient second order minimization technique (ESM) [15].

However, most methods assume that the luminance remains constant between two successive frames, which is not true. Indeed, most surfaces are not Lambertian and lighting conditions are mostly variable during an image sequence. Hager and Belhumeur [16] propose to acquire an image database of the scene under several illuminations and to use these data to improve the tracking. Although efficient, this method requires a prior learning step, which can be seen as too restrictive. Alternatively, one can either use a simple local photometric normalization as in [17] or compute a photometric model which properly fits the luminance variations in small areas of the image as [18, 19, 20, 21] for optical flow, [22] for object recognition, or [23] for indexing. In [24], the tracker compensates for affine illumination changes. More recently in [25], the authors compute arbitrary illumination changes on a large planar area by using an ESM algorithm. The main difficulty of the illumination compensation is to balance the trade-off between complexity and adequacy with the real illumination changes. Moreover, these models are based on several assumptions which have not been clearly defined yet.

Our first contribution is to clearly explain the modeling of luminance changes due to lighting and geometry, by analyzing some widely used specu-

lar reflection models [26, 27]. We particularly focus on two local approximations: the first one is well suited for specular highlights and lighting changes when small areas are concerned; the second one remains appropriate for larger windows.

The second contribution of the paper consists in studying the validity of the proposed models, which is related to the object parameters and the acquisition setup. Note that the surfaces are considered to be differentiable, which is a reasonable assumption for: 1) small surfaces such as points, 2) industrial applications where manufactured objects and natural surfaces have to be manipulated.

Finally, we propose a KLT-like tracker. Compared to most efficient KLT-like trackers, it leads to the best results in terms of the number of points correctly tracked, residuals and location errors.

This article is organized as follows. Section 2 focuses on the general modeling of luminance changes, especially in the case of specular reflections and lighting variations. Then, Section 3 deals with the local illumination models used for temporal matching, and then details the two photometric models proposed. The theoretical validity of the photometric models is studied in Section 4 by considering several specific configurations on the viewing geometry and the surface properties. Section 5 details some of the existing trackers, regarding the illumination model on which they are based, and explains the proposed methods. The relevance of the present work is validated through experimental results in Section 6.

2. Modeling of luminance changes

The description of the luminance is detailed hereafter, it is based upon simplified physical models largely used in image analysis. For image synthesis, some more elaborated BRDF are considered. This study focuses on the luminance changes occurring between two successive images of the same scene.

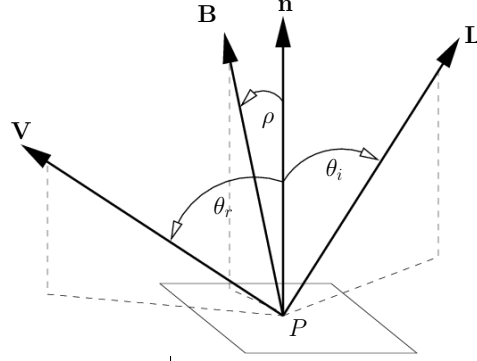
A uniform lighting and a point light source are considered, which is well suited for our targeted applications in industrial contexts or outdoor scenes where the is considered as a point light source. More generally, each point in the scene is assumed to be locally lighted by a dominant point light source.

First, let us introduce our notations (see Figure 1). Let P be a point of the object being observed, and \mathbf{V} and \mathbf{L} the viewing and lighting directions respectively, which form the angles θ_r and θ_i with the normal \mathbf{n} in P . \mathbf{B} is the bisecting line between \mathbf{V} and \mathbf{L} , and forms an angle ρ with \mathbf{n} . f and f' are two successive images of the same scene. P projects in image f into p of coordinates (x_p, y_p) and into p' of coordinates (x'_p, y'_p) in the image f' after a relative motion between the camera, the scene and the light sources. More generally, the prime symbol refers to the parameters in f' .

2.1. The luminance in the CCD plane

The luminance $f(p)$ results from the integration of the radiance $\mathcal{L}_P(\lambda)$ emitted by P w.r.t the wavelength λ . Since the radiance is the product of the illuminant spectrum $\mathcal{E}_P(\lambda)$ with the material reflectance $\mathcal{R}_P(\lambda)$ then

$$f(p) = k_c \int_{\lambda_{\min}}^{\lambda_{\max}} \mathcal{S}(\lambda) \mathcal{L}_P(\lambda) d\lambda = k_c \int_{\lambda_{\min}}^{\lambda_{\max}} \mathcal{S}(\lambda) \mathcal{E}_P(\lambda) \mathcal{R}_P(\lambda) d\lambda. \quad (1)$$



P	a physical point in the scene	f	initial image
\mathbf{V}	viewing direction	f'	image with photometric and geometric changes
\mathbf{L}	lighting direction	p	projection of P in f
\mathbf{B}	bisecting line between \mathbf{V} and \mathbf{L}	p'	projection of P in f'
\mathbf{n}	normal vector in P	(x_p, y_p)	coordinates of p in f
θ_r	viewing angle (\mathbf{n}, \mathbf{V})	(x'_p, y'_p)	coordinates of p' in f'
θ_i	incident angle (\mathbf{n}, \mathbf{L})	$\boldsymbol{\mu}$	motion model
ρ	specular angle (\mathbf{n}, \mathbf{B})	$\delta(\cdot)$	motion function parameterized by $\boldsymbol{\mu}$
			such that $\delta(p, \boldsymbol{\mu}) = p'$

Figure 1: Notations involved in the reflection description.

where $\mathcal{S}(\lambda)$ describes the spectral sensitivity of the sensor and k_c is a scalar which does not depend on the wavelength λ but only on the optical geometry such as the focal distance and the aperture.

The reflectance is generally assumed to be diffuse or Lambertian [28], i.e it does not depend on θ_r and it is a function of the *body reflection* called $\mathcal{R}_P^b(\lambda)$.

However, most surfaces reflect light in a specular manner, not only in a diffuse way. For example, the Phong-Blinn model [26, 29] describes the

reflectance of specular surfaces in the following heuristic way, for $\theta_i(P) \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$:

$$\mathcal{R}_P(\lambda) = k_d(P)\mathcal{R}_P^b(\lambda) \cos \theta_i(P) + k_s(P) \cos^r(\rho(P)) + k_a(P) \quad (2)$$

The first term is the Lambertian contribution while the second one is the specular component, where the scalar r is inversely proportional to the roughness of the surface and k_s , k_a are the specular and ambient gains in P . Although empirical, this model is largely used, because of its simplicity, and its validity for various types of materials, whether they are rough or smooth. Note that the Torrance-Sparrow model [27], is also widely used. In both models, the specular term reaches its maximum value for $\rho(P) = 0$, that is when \mathbf{B} coincides with \mathbf{n} . In the remainder of the paper, this specular term in P is noted $h(P)$.

2.2. The modeling of the luminance in the image

Consider $\mathcal{M}_P(\lambda) = \mathcal{S}(\lambda)\mathcal{E}_P(\lambda)$ in (1). When the sensor has a linear response, $\mathcal{M}_P(\lambda)$ can be expressed as the product of a global gain $k_m(P)$ with a spectrum curve $\mathcal{M}(\lambda)$, and (1) becomes:

$$f(p) = k_c k_m(P) \int_{\lambda_{min}}^{\lambda_{max}} \mathcal{M}(\lambda) \mathcal{R}_P(\lambda) d\lambda. \quad (3)$$

Considering (1) and (2), the luminance f can be modeled as a sum of three terms¹:

$$f(p) = K_d(p)a(p) \cos \theta_i(p) + h(p) + K_a(p) \quad (4)$$

¹Note P has been replaced by its projection into the sensor p .

where the term h refers to the specular reflection function (see 2.1) and includes a gain $K_s(p)$. Note also that $K_d(p) = k_c k_m(p) k_d(p)$, $K_s(p)$ and $K_a(p)$ are the integration values respectively of $k_c k_m(p) k_s(P) \mathcal{M}(\lambda)$ and $k_c k_m(p) k_a(P) \mathcal{M}(\lambda)$ w.r.t the wavelengths.

The term noted $a(p) = \int_{\lambda_{min}}^{\lambda_{max}} \mathcal{M}(\lambda) \mathcal{R}_P^b(\lambda) d\lambda$ depends on the *albedo* $\mathcal{R}_P^b(\lambda)$, thus on the intrinsic property of the material.

2.3. The luminance changes between two images of a sequence

Two categories of illumination variations can occur between two successive frames, either they are due to a sole motion of the camera w.r.t P , or they are related to lighting changes (intensity change, or relative motion between the source and the object).

Motion of the camera. Some specular reflections can occur due to a simple motion of the camera with respect to the surface, leading to the displacement of p towards p' . The incident angle θ_i remains constant over time $\theta'_i(p') = \theta_i(p)$ and if no lighting change occurs, K_d and K_a are also constant. Similarly, $a'(p') = a(p)$ since it is an intrinsic property of the material. However, the specular component h can vary since it depends on the viewing direction through the angle ρ . Thus, the luminance f' becomes

$$f'(p') = K_d a(p) \cos \theta_i(p) + h'(p') + K_a \quad (5)$$

By subtracting (4) from (5), it yields:

$$f'(p') = f(p) + \gamma(p) \text{ with } \gamma(p) = h'(p') - h(p). \quad (6)$$

Lighting changes. Now, some lighting changes ΔK_a , ΔK_d occur on K_a and K_d respectively. They stem from a shift of the camera gain, a variation

of lighting intensity or shadows. Moreover, θ_i changes in p according to a function $\Delta\theta_i$, and so the specular term $h'(p')$. The luminance f' is then expressed as

$$f'(p') = K'_d(p') a(p) \cos \theta'_i(p') + h'(p') + K'_a \quad (7)$$

with:

$$\begin{cases} K'_d(p') &= K_d(p) + \Delta K_d(p) \\ \theta'_i(p') &= \theta_i(p) + \Delta\theta_i(p) \\ K'_a &= K_a + \Delta K_a. \end{cases} \quad (8)$$

The specular term $h'(p')$ includes the intensity change of the specular coefficient K_s . Therefore, by using (4) and (7), the relationship between two images of the same scene can be described by two different expressions:

1) First, it can be analogous to (6) by subtracting (4) from (7), where the function γ is given by the following relationship:

$$\begin{aligned} \gamma(p) &= a(p)(K'_d(p') \cos(\theta_i(p) + \Delta\theta_i(p)) - K_d(p) \cos \theta_i(p)) + \\ &\quad h'(p') - h(p) + \Delta K_a \end{aligned} \quad (9)$$

Here, $\gamma(p)$ depends on $a(p)$ and thus on the *albedo*.

2) Second, it can be expressed by:

$$f'(p') = \alpha(p)f(p) + \beta(p) \quad (10)$$

where:

$$\begin{cases} \alpha(p) &= \frac{(K_d(p') + \Delta K_d(p')) \cos(\theta_i(p) + \Delta\theta_i(p))}{K_d(p) \cos \theta_i(p)} \\ \beta(p) &= -(h(p) + K_a)\alpha(p) + h'(p') + K_a + \Delta K_a. \end{cases} \quad (11)$$

Note that neither $\alpha(p)$ nor $\beta(p)$ depend on $a(p)$, but only on the geometry. Due to the large number of parameters, among which the material properties

(the roughness of the surface by the means of the specular terms), $\alpha(p)$ and $\beta(p)$ are not easy to compute and therefore have to be approximated.

3. Local modeling of illumination changes in the image

Generally speaking, the simplified photometric models rely on the modeling of luminance changes in local areas of the image. Therefore, they are available for image matching or feature point tracking when surfaces are assumed to be differentiable. Starting from (10) and (11), this section studies the assumptions on which these models are based. Let \mathcal{W} be a window of interest centered in p and m a neighbor point belonging to \mathcal{W} .

3.1. The luminance constancy

From the radiance models of Section 2.1, the luminance constancy [9] is true only for Lambertian objects under constant lighting. In that case:

$$f'(m') = f(m) \text{ for any } m \in \mathcal{W}. \quad (12)$$

3.2. The affine model

The affine model assumes that the photometric variations are locally constant so that $\alpha(p) = \alpha$ and $\beta(p) = \beta$ in (10) leading to:

$$f'(m') = \alpha f(m) + \beta. \quad (13)$$

Remark 1: the affine photometric model, given by (13) and the photometric normalization ([17] for example) are similar [30]. Let be μ_f and σ_f (respectively $\mu_{f'}$ and $\sigma_{f'}$) the average and standard deviation of f (respectively f') computed in \mathcal{W} . The normalization assumes the following

relationship:

$$\frac{f'(m') - \mu_{f'}}{\sigma_{f'}} = \frac{f(m) - \mu_f}{\sigma_f} \quad (14)$$

Therefore:

$$\begin{cases} \alpha = \frac{\sigma_{f'}}{\sigma_f} \\ \beta = \mu_{f'} - \frac{\sigma_{f'}\mu_f}{\sigma_f} \end{cases} \quad (15)$$

Remark 2: The normalization (14) may introduce some noise for low σ i.e for high luminance saturation or homogeneity in \mathcal{W} .

According to (11), the affine model (13) assumes that:

- 1) the surface is locally planar. Indeed, the incident angles θ_i and $\Delta\theta_i$ do not vary spatially in \mathcal{W} , i.e. when \mathbf{n} is the same at each point of \mathcal{W} ;
- 2) both objects and lighting are motionless (θ_i constant);
- 3) the surface is Lambertian. Indeed, h' and h are constant in \mathcal{W} . According to (2), this is true when both the ρ and the roughness r are constant in \mathcal{W} , i.e. when for all m in \mathcal{W} , $h(m) = h'(m') = 0$ (no specular term).

3.3. Some models suited for specular highlights occurrence and lighting changes

The above assumptions are incorrect for non-planar surfaces, for which all the angles vary in \mathcal{W} . This section proposes two models which are less restrictive. The first one is available for small windows of interest.

3.3.1. A photometric model for small areas

As shown by Section 2, if only the camera moves, the luminance variations between two frames are properly described by (6), otherwise by (9).

According to (2), γ is not constant in \mathcal{W} since it depends on the angles and therefore on the normal \mathbf{n} at each point of \mathcal{W} . It also depends on the γ

can be correctly approximated at first order on \mathcal{W} by a \mathcal{C}^2 function², noted γ_{mod} ,

which can be expanded in Taylor series, centered at a point m with coordinates (x, y) , in the neighborhood of p :

$$\gamma(m) \simeq \gamma_{mod}(p) + \left. \frac{\partial \gamma_{mod}}{\partial x} \right|_p (x - x_p) + \left. \frac{\partial \gamma_{mod}}{\partial y} \right|_p (y - y_p). \quad (16)$$

Let be $\gamma_1 = \left. \frac{\partial \gamma_{mod}}{\partial x} \right|_p$, $\gamma_2 = \left. \frac{\partial \gamma_{mod}}{\partial y} \right|_p$ and $\gamma_3 = \gamma_{mod}(p)$, $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \gamma_3)$ and $\mathbf{u} = (x - x_p, y - y_p, 1)$. By injecting (16) in (6) it yields:

$$f'(m') = f(m) + \boldsymbol{\gamma}^\top \mathbf{u} \quad (17)$$

This model is appropriate to model the photometric variations due to:

- 1) the motion of the camera, when there is no relative motion between the light source and the object. The surface projected onto \mathcal{W} is not mandatory planar but its normal has to vary smoothly, and K_s and r as well (see (2)). In that context, the model (17) is well suited to compensate for specular highlights.
- 2) lighting changes, when the *albedo* and the normal \mathbf{n} vary smoothly (see equation (9)). The approximation of $a(m)$ by a first order polynomial becomes more and more crude for large and textured surfaces.

3.3.2. A photometric model for large areas

According to (11), α depends on θ_i , which varies on singular surfaces, especially on wide \mathcal{W} . Likewise, β depends on the non-uniform specular highlights since they depend on \mathbf{n} and ρ . However, it is reasonable to assume

²It implies that \mathbf{n} varies smoothly in \mathcal{W} , *i.e* the surface is regular.

that they are continuous and differentiable in each point m when \mathbf{n} , h and r vary smoothly in \mathcal{W} . Then, α and β can be expanded in Taylor series around p . By writing $\boldsymbol{\alpha} = \left(\frac{\partial \alpha}{\partial x} \Big|_p, \frac{\partial \alpha}{\partial y} \Big|_p, \alpha(p) \right)$ and $\boldsymbol{\beta} = \left(\frac{\partial \beta}{\partial x} \Big|_p, \frac{\partial \beta}{\partial y} \Big|_p, \beta(p) \right)$, it yields

$$f'(m') = \boldsymbol{\alpha}^\top \mathbf{u} f(m) + \boldsymbol{\beta}^\top \mathbf{u} \quad (18)$$

This model assumes that the photometric changes can be non-uniform on a same \mathcal{W} . It is suited locally for both highlights and lighting changes, only when the surface is smooth. The validity of this model is studied more thoroughly in the next section.

4. Study on the validity of the photometric model

Consider a planar object³ viewed under one point light source at a known location. We compute the real photometric changes (α and β given by (11)) resulting from a change of the pose of the camera and the light source with regard to the object.

Second, we achieve a local approximation of these photometric functions by computing the second order Taylor series of α and β . The first order approximation of (18) is suited when at least its coefficients of second order are null or almost null. So, this study consists in finding these specific configurations⁴.

³Obviously, the photometric model will be less adequate for non-planar surfaces [31].

⁴To be comprehensive, the third and upper orders should be analyzed, but this fastidious analysis would have been detrimental to the clarity and concision of the paper.

4.1. Modeling of the scene geometry

Consider a frame \mathcal{F}_c linked to the camera. A point P with coordinates (X_p, Y_p, Z_p) is located at the center of a region of interest on the object. Let us also consider a point M , with coordinates (X, Y, Z) , located in the neighborhood of P . We assume that the surface in P can be described as a \mathcal{C}^2 function leading to the following approximation of the depth in M , assumed to be valid in each point of \mathcal{W} :

$$Z = Z_P + D_X(X - X_P) + D_Y(Y - Y_P), \quad (19)$$

where the first derivatives of the surface D_X, D_Y describe the orientation between the tangent plane of the surface at P and the CCD plane: $D_X = \frac{\partial Z}{\partial X}|_P$ and $D_Y = \frac{\partial Z}{\partial Y}|_P$. The normal vector in P is $\mathbf{n} = (\frac{\partial Z}{\partial X}, \frac{\partial Z}{\partial Y}, -1)$.

Given $S = (S_x, S_y, S_z)$ (in the frame \mathcal{F}_c) the location of the light source, let $\mathbf{L} = (X - S_x, Y - S_y, Z)$ be the vector linking the lighting source S to the point M . Then, $\cos \theta_i = \mathbf{L} \cdot \mathbf{n}$. By perspective projection and by using Z given by (19), the real parameters α and β given by (11) can be expressed with respect to the pixels coordinates m . Therefore some approximations and Taylor series expansions are achieved according to the acquisition configurations.

First of all, we study the validity of the approximation of α (section 3.3.2) by a first order polynomial. In order to simplify this study, we focus on small windows of interest \mathcal{W} which are located near the optical axis of the camera.

4.2. Validity of the approximation of α

Let be $u = x - x_P$ and $v = y - y_P$ and consider the approximation of α (see equation (11)) at second order:

$$\alpha(m) = \alpha_1 u + \alpha_2 v + \alpha_3 + \alpha_4 u^2 + \alpha_5 v^2 + \alpha_6 uv \quad (20)$$

The lighting conditions for which $(\alpha_4, \alpha_5, \alpha_6)$ can be neglected are those for which the photometric model (18) better fits the illumination changes.

Only the direct light source is moved from its initial position S with a small motion $\mathbf{dS} = (dS_X, dS_Y, dS_Z)$. Indeed, a single motion of the light source impacts both α and β . Assuming a small \mathbf{dS} , the coefficients α_i can be expanded in Taylor series around \mathbf{dS} . Let us focus on the expression of α_4 at first order, since the two other coefficients $\alpha_5 = \alpha_6$ do not provide additional information. Three different configurations are considered.

1. The lighting vector \mathbf{L} coincides with \mathbf{n} : $\mathbf{L} = \tau \mathbf{n}$ with τ constant:

$$\alpha_4 = \alpha_5 = \alpha_6 = 0 \quad (21)$$

2. The light source is close to the camera ($S = O$ and $\mathbf{V} = \mathbf{L}$):

$$\alpha_4 = -\frac{1}{Z_P} (2dS_Z + 2D_X dS_X) \quad (22)$$

3. The lighting source is close to the surface ($S = (X_p, Y_p, Z_p - \epsilon)$ and Taylor expansion around $\epsilon = 0$).

$$\alpha_4 = -\frac{2Z_P(Z_P + D_X^2\epsilon - Z_P D_X^2)}{\epsilon^3} dS_Z + \frac{2Z_P(D_X(\epsilon - 1))}{\epsilon^3} dS_X \quad (23)$$

Thus, our approximation of α is particularly well adapted when \mathbf{L} coincides with \mathbf{n} at the considered point (see (21)) and when the tangent plane of the

surface at P is parallel to the sensor plane *i.e.* when \mathbf{V} coincides with \mathbf{n} (indeed, $D_X = 0$ in (22) and (23)). The model is also well adapted when the depth of camera and light source are high (high Z_P in (22)). When the lighting source is close to the surface, the model is correct when the camera is even closer to the surface (low Z_P in (23)). Note also that a backward or forward motion of the light source influences the coefficient values, since dS_Z appears in (22) and (23).

4.3. Validity of the approximation of β

In order to study the validity of the modeling of β (expressed by (11)), it is necessary to consider the specular highlights and therefore the material properties. For this purpose, we use the Phong model (2). We assume a static object under constant lighting (K_a and K_d), so that $\alpha(m) = 1$. Consequently, β is equivalent to the function γ in (6). Thus, we study the validity of the following expression: $\beta(m') = h'(m) - h(m)$. β is expanded in Taylor series at second order around p :

$$\beta(m) = \beta_1 x + \beta_2 y + \beta_3 + \beta_4 x^2 + \beta_5 y^2 + \beta_6 xy \quad (24)$$

where the coefficients β_i depend on the geometry parameters explained in section 4.1. Since h reaches its maximum when ρ is null, it is interesting to study the validity of the photometric models in this configuration, *i.e.* \mathbf{n} coincide with \mathbf{B} (see Figure 1). As previously, we assume a small motion of the light source $d\mathbf{S}$ and achieve a Taylor series expansion of (24) around $d\mathbf{S}$. Some particular configurations of the scene geometry are studied in order to reach some simple conclusions.

1. $\mathbf{L} = \mathbf{V} = \mathbf{n}$, therefore $D_X = D_Y = 0$ and $\theta_i = 0$. Since the expression of β_4 is far too complicated to deduce any useful information, let us focus on two particular configurations:

(a) The light source is close to the surface ($Z_S = Z_P$), then:

$$\beta_4 = -r \left(\frac{1}{Z_P} \right) dSZ \quad (25)$$

(b) The sensor is close to P (Z_P low, and Taylor series around $Z_P = 0$), then:

$$\beta_4 = \beta_5 = \beta_6 = 0 \quad (26)$$

2. Small orientation of the surface with regard to the sensor plane (small D_X and D_Y and Taylor series around $D_X = D_Y = 0$, $\mathbf{n} = \mathbf{B}$).

(a) The light source is located near the surface $Z_S = Z_P$

$$\eta_4 = -\frac{r}{Z_P} \left(\frac{D_X}{4}(3r+7)dS_X + \frac{D_X}{4}(r+1)dS_Y + dS_Z \right) \quad (27)$$

(b) The sensor is close to the surface

$$\beta_4 = \beta_5 = \beta_6 = 0 \quad (28)$$

According to (26) and (28), our photometric approximation is appropriate when the sensor is close to the surface. Otherwise, a forward (or a backward) motion dS_Z of the light source w.r.t the surface induces variations of the β_i . A motion along the Z axis has less impact when the sensor is far from the surface (large Z_P) and when the surface is rough (low r in (25) or (27)).

In those conditions, the variations of the specular highlight draw up a plane on the window of interest \mathcal{W} , which can properly be compensated by the proposed illumination model.

4.4. Summary

Table 1 provides an overview of the configurations for which the proposed photometric model is adapted (+) or not (-), or when the configuration has no influence (=). Note that the affine photometric model (see section 3.2) relies on more restrictive assumptions, since α_2 , α_3 , β_2 , β_3 have also to be null in (24) or (20). As said in section 3.2 page 10, the affine model is suited for Lambertian surfaces and is not adapted for a motion of a light source w.r.t to the surface.

The photometric models can be used for matching in order to improve some higher level procedures [21]. In this paper, we improve the robustness of KLT-like tracking.

5. Feature points tracking algorithms

Tracking features in a robust and accurate way along an image sequence is still an open problem, which explains the large variety of existing methods. The relative motion between the camera and the scene induces a geometric deformation described by the function δ which links m' to m such that $\delta(m, \boldsymbol{\mu}) = m'$ for all $m \in \mathcal{W}$, according to a parameterization described by the vector $\boldsymbol{\mu}$. The feature point tracking consists in computing $\boldsymbol{\mu}$ according to a photometric model. We will show how to compute $\boldsymbol{\mu}$ for the photometric models given in Section 3.

5.1. Commonly used tracking methods

5.1.1. The classical approach

The classical KLT technique (for Kanade-Tomasi-Lucas tracker [7, 8]) assumes a Lambertian object, thus a perfect conservation of luminance at a

Table 1: Overview of the results about the validity of the approximations of α and β by a Taylor series expansion at first order. +: good approximation. -: bad approximation. =: there is no influence on the validity.

Configuration	α	β
Lighting vector coincides with the normal	+	+
Viewing vector coincides with the normal	+	+
Rough surface	=	+
Sensor close to the surface and lighting source far from the surface	+	+
Motion of the lighting source along the optical axis	-	-

point during the sequence (see (12)), so $f(m) = f'(\delta(m, \boldsymbol{\mu}))$. However, as seen in Section 2, the luminance constancy assumption is not true. Besides, $\delta(m, \boldsymbol{\mu})$ is also an approximation. Thus, it is more judicious to minimize the following criterion:

$$\epsilon_1(\boldsymbol{\mu}) = \sum_{m \in \mathcal{W}} (f(m) - f'(\delta(m, \boldsymbol{\mu})))^2 \quad (29)$$

In order to obtain $\boldsymbol{\mu}$, we suppose that $\boldsymbol{\mu} = \hat{\boldsymbol{\mu}} + \Delta\boldsymbol{\mu}$, where $\Delta\boldsymbol{\mu}$ expresses a small variation around an estimation $\hat{\boldsymbol{\mu}}$ of $\boldsymbol{\mu}$. When $\hat{\boldsymbol{\mu}}$ is unknown, it is initialized as $\hat{\boldsymbol{\mu}} = 0$. In those conditions, $f'(\delta(m, \boldsymbol{\mu}))$ can be expanded in a first order Taylor series around $\hat{\boldsymbol{\mu}}$:

$$f'(\delta(m, \boldsymbol{\mu})) = f'(\delta(m, \hat{\boldsymbol{\mu}})) + \nabla f'^\top(\delta(m, \hat{\boldsymbol{\mu}})) J_{\delta}^{\hat{\boldsymbol{\mu}}} \Delta\boldsymbol{\mu} \quad (30)$$

where $J_{\delta}^{\hat{\boldsymbol{\mu}}}$ is the Jacobian of δ according to $\boldsymbol{\mu}$, expressed in $\hat{\boldsymbol{\mu}}$. Substituting (30) into (29) leads to a linear system in $\Delta\boldsymbol{\mu}$, which can be solved iteratively:

$$\left(\sum_{m \in W} \mathbf{v}_c \mathbf{v}_c^T \right) \Delta \boldsymbol{\mu} = \sum_{m \in W} (f(m) - f'(\delta(m, \hat{\boldsymbol{\mu}}))) \mathbf{v}_c \quad (31)$$

with $\mathbf{v}_c = (J_{\delta} \hat{\boldsymbol{\mu}})^T \nabla f'(\delta(m, \hat{\boldsymbol{\mu}}))$. When considering an affine motion model, \mathbf{v}_c is the vector defined by $\mathbf{v}_c = (f'_x, f'_y, x f'_x, x f'_y, y f'_x, y f'_y)^T$, where f'_x and f'_y are the derivatives of f' with respect to x and y respectively.

5.1.2. Tracking methods robust to affine photometric changes

These approaches are based on the photometric model described in section 3.2. Therefore, the minimization function becomes:

$$\epsilon_2(\boldsymbol{\mu}, \alpha, \beta) = \sum_{m \in W} (\alpha f(m) + \beta - f'(\delta(m, \boldsymbol{\mu})))^2, \quad (32)$$

where α and β refer to the parameters of the affine model (13). There are two ways to obtain α and β , either by computing them outside the minimization loop (32) [17] or by computing them simultaneously with $\boldsymbol{\mu}$.

The photometric normalization. The tracking technique consists in computing $\boldsymbol{\mu}$ as in Section 5.1.1 since α and β are constant. We have to solve:

$$\left(\sum_{m \in W} \mathbf{v}_c \mathbf{v}_c^T \right) \Delta \boldsymbol{\mu} = \sum_{m \in W} (\alpha f(m) + \beta - f'(\delta(m, \hat{\boldsymbol{\mu}}))) \mathbf{v}_c \quad (33)$$

where α and β are computed using (15).

Estimation of α and β : the Jin's technique. In [24], the authors propose to estimate the contrast α and intensity β simultaneously with the motion model.

Let $\boldsymbol{\nu}$ be the vector of photometric variations $\boldsymbol{\nu} = (\alpha, \beta)$, and \mathbf{d} the concatenation of $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$. As previously, we assume a small variation $\Delta \mathbf{d} =$

$(\Delta\boldsymbol{\mu}, \Delta\boldsymbol{\nu})$ of \mathbf{d} around its estimate $\hat{\mathbf{d}}$ so that $\mathbf{d} = \hat{\mathbf{d}} + \Delta\mathbf{d}$. Thus, by using (30) and introducing $\hat{\boldsymbol{\nu}}$, (32) becomes

$$\left(\sum_{m \in \mathcal{W}} \mathbf{v}_s \mathbf{v}_s^\top \right) \Delta\mathbf{d} = \sum_{m \in \mathcal{W}} \left(\hat{\alpha} f(m) + \hat{\beta} - f'(\delta(m, \hat{\boldsymbol{\mu}})) \right) \mathbf{v}_s \quad (34)$$

where $\mathbf{v}_s = (-\mathbf{v}_c, f(m), 1)^\top$.

Refer to Appendix A which compares the conditioning of the matrices used in the trackers detailed in the paper. The matrix $\sum_{m \in \mathcal{W}} \mathbf{v}_s \mathbf{v}_s^\top$ is ill-conditioned and cannot be inverted easily, therefore a preconditioning, adapted to the image, is required.

The next section proposes two tracking procedures to account for non-uniform photometric changes.

5.2. Proposed tracking procedures

5.2.1. A tracking method robust to specular highlights

The first tracking method is based on the illumination model given by (17). The following criterion has to be minimized:

$$\epsilon_3(\boldsymbol{\mu}, \boldsymbol{\gamma}) = \sum_{m \in \mathcal{W}} \left(f(m) - f'(\delta(m, \boldsymbol{\mu})) + \boldsymbol{\gamma}^\top \mathbf{u} \right)^2 \quad (35)$$

Let $\mathbf{d} = (\boldsymbol{\mu}, \boldsymbol{\gamma})$ and further assume a small motion $\Delta\mathbf{d} = (\Delta\boldsymbol{\mu}, \Delta\boldsymbol{\gamma})$ around an estimate $\hat{\mathbf{d}}$ of \mathbf{d} . Similarly to the classical method of Section 5.1.1, $\Delta\mathbf{d}$ is computed by solving the following linear system:

$$\left(\sum_{m \in \mathcal{W}} \mathbf{v}_p \mathbf{v}_p^\top \right) \Delta\mathbf{d} = \sum_{m \in \mathcal{W}} \left(f(m) - f'(\delta(m, \hat{\boldsymbol{\mu}})) + \hat{\boldsymbol{\gamma}}^\top \mathbf{u} \right) \mathbf{v}_p \quad (36)$$

where $\mathbf{v}_p = (-\mathbf{v}_c, \mathbf{u})^\top$. Unlike the previous tracker, the matrix $(\sum_{m \in \mathcal{W}} \mathbf{v}_p \mathbf{v}_p^\top)$ is well conditioned (see Appendix A).

According to the assumptions of the photometric model (17), this tracker is more appropriate to compensate for specular highlights. For small \mathcal{W} , it can also compensate for lighting changes when (9) can be approximated by a first order Taylor series expansion. For large windows, the next method is more appropriate.

5.2.2. A tracking method robust to specular highlights and lighting changes

Section 3.3.2 has detailed a comprehensive photometric model which compensates for spatial variations of specular highlights and lighting changes. The parameters $\boldsymbol{\mu}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are computed by minimizing the following criterion

$$\epsilon_4(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{m \in \mathcal{W}} (\boldsymbol{\alpha}^\top \mathbf{u} f(m) - f'(\delta(m, \boldsymbol{\mu})) + \boldsymbol{\beta}^\top \mathbf{u})^2 \quad (37)$$

The system can be linearized as in section 5.2.1, with $\mathbf{d} = (\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\beta})$. Thus, the tracking consists in solving the following system:

$$\left(\sum_{m \in \mathcal{W}} \mathbf{v}_m \mathbf{v}_m^\top \right) \Delta \mathbf{d} = \sum_{m \in \mathcal{W}} \left(\hat{\boldsymbol{\alpha}}^\top \mathbf{u} f(m) - f'(\delta(m, \hat{\boldsymbol{\mu}})) + \hat{\boldsymbol{\beta}}^\top \mathbf{u} \right) \mathbf{v}_m \quad (38)$$

where $\mathbf{v}_m = (-\mathbf{v}_c, f(m)\mathbf{u}, \mathbf{u})^\top$. Since the values of \mathbf{v}_m are often dissimilar, the matrix $\sum_{m \in \mathcal{W}} \mathbf{v}_m \mathbf{v}_m^\top$ can be ill-conditioned (see Appendix A).

Moreover, the number of parameters is quite large. Indeed, by using an affine motion model, twelve parameters have to be computed. Obviously, the use of too small windows of interest may alter the accuracy of both photometric and motion estimations.

6. Validation and experimental results

This section compares the previous trackers through several sequences showing geometric and photometric changes simultaneously. First of all, we

detail the experimental setup, the notations and the outlier rejection strategy. The first experiments are carried out on sequences for which a ground-truth is available in order to evaluate the methods as well as the outliers rejection procedure. Then, experiments are performed on *real sequences*, without ground-truth.

6.1. Experimental setup and evaluation with ground-truth

Tracking methods. The following notations are used:

C: classical tracking (section 5.1.1) for which $f'(\delta(m, \boldsymbol{\mu})) = f(m)$

N: photometric normalization (section 5.1.2): $f'(\delta(m, \boldsymbol{\mu})) = \lambda f(m) + \eta$

J: Jin *et al.* method (section 5.1.2): $f'(\delta(m, \boldsymbol{\mu})) = \lambda f(m) + \eta$

P_3 : tracker with three parameters (section 5.2.1): $f'(\delta(m, \boldsymbol{\mu})) = f(m) + \mathbf{u}^\top \boldsymbol{\gamma}$

P_6 : tracker with six parameters (section 5.2.2): $f'(\delta(m, \boldsymbol{\mu})) = \boldsymbol{\alpha}^\top \mathbf{u} f(m) + \boldsymbol{\beta}^\top \mathbf{u}$

Size of the window of interest. Usually, the choice of the size $\mathcal{N} \times \mathcal{N}$ of the square window \mathcal{W} is based on a trade-off between robustness to noise, computation times and reliability of the assumptions on which the tracker is based. Here, we consider some sizes ranging from $\mathcal{N} = 9$ to $\mathcal{N} = 35$.

Point rejection and comparison criteria. The points are selected in the first frame of the sequence by the Harris detector [5] and an affine motion model is computed between the first frame and the current frame.

The rejection process and its validation. The tracker has to constantly check whether the points are correctly tracked and are reliable⁵. To this end, most techniques analyze the residuals [24, 16]. In our work, a point is rejected as soon as its residuals become higher than a threshold,

⁵They can be lost because of noise, disappearance or bad convergence.

$\mathcal{S}_{conv} = \mathcal{N}^2 E_{ave}^2$, where E_{ave} is the tolerated luminance variation for each point in \mathcal{W} between f and its model. We choose $E_{ave} = 15$. For each sequence, the following evaluation criteria are considered:

1) *The robustness* of the tracking, *i.e.* the number of points correctly tracked during the whole sequence.

2) *The temporal evolution of the mean convergence residuals*: $(1/n_k) \sum_{i=1}^{n_k} \epsilon^i$, where n_k is the number of points currently tracked by the considered method k . These two first criteria have to be considered jointly since the most satisfactory method is the one which tracks a large number of points with low residuals.

3) *The location errors*. When ground-truth is available, the mean location error is computed on all the points that are correctly tracked by the technique. Note that for a same location error, the best technique is the one which tracks the larger number of points.

Computation of the ground-truth. For planar surfaces, four blobs, easy to segment, are put on the object in order to compute the homography transform from the initial frame to the current one. For non planar surfaces, the pose is computed between the camera and the object, which is static (see [30] for the implementation details). In each case, the error is given by the euclidean distance between the theoretical points and the location provided by the tracker.

Sequences. In the two sequences used, there is no lighting change but the motion yields specular variations. Glossy paper has been chosen to cover the objects. The sequence *Planar* (Fig.2(a)) shows a planar surface of size 1×1 m. The camera is fixed at $Z_P = 4$ m and two light sources are located

at $S_z \simeq 2$ m. The object is moved by an operator. The sequence *Cylinder* (Fig.2(b)) shows a cylinder of radius 7 cm. The object and light sources are static, the camera is moving and $Z_P \simeq 1$ m and $S_z > Z_P$.

Results.

1) *Robustness.* Table 2(a) and 2(b) page 26 show the percentage of correct points. For small windows, P_3 tracks the highest number of points. In contrast, for larger \mathcal{W} , P_6 is the most competitive method: on the planar surface it correctly tracks around twice the number of points. Despite its bad conditioning (see Appendix A), it outperforms J or N by better compensating the specular changes.

2) *Location errors.* Fig. 3 and 4 compare the mean location error computed on the correct points⁶. The location errors are low: around 1 pixel at the beginning of the sequence and less than 1 pixel when outliers points are rejected. The strategy for outliers rejection, based on the residuals, is therefore efficient to maintain a good accuracy of the tracking.

6.2. Experiments on real sequences

In the next sequences, the camera is moving and the scene is motionless:

1) Sequence *Cylinder2* (150 images) (see Fig. 5a) shows a specular cylinder. The lighting conditions do not change but the motion of the camera causes some specular highlight variations on the object surface.

2) Sequences *Planar object* (Fig. 5b) and *Marylin* (Fig. 5c) show several textured objects consisting of several materials (glossy paper, ceramic, metal, cardboard, glass) under an ambient lighting (daylight and fluorescent lamps

⁶The method C is not taken into account since too many points are lost.

Table 2: *Lab sequences*. Percentage of points which have been correctly tracked during the sequence (occluded points or points going out of the image are not taken into account). *C*: classical tracking. *N*: tracking with photometric normalization. *J*: Jin *et al.* method. *P*₃: tracker with three photometric parameters. *P*₆: tracker with six parameters.

(a) PLANAR SURFACE.				(b) CYLINDER.				
\mathcal{N}	9	15	25	\mathcal{N}	9	15	25	35
<i>C</i>	6	9	6	<i>C</i>	8	24	14	10
<i>N</i>	29	32	32	<i>N</i>	2	30	36	40
<i>J</i>	23	37	43	<i>J</i>	0	8	28	34
<i>P</i> ₃	34	34	37	<i>P</i> ₃	10	40	70	62
<i>P</i> ₆	0	46	69	<i>P</i> ₆	0	0	34	68

located on the ceiling) and a direct light source. In the sequence *Planar object*, the intensity of the direct lighting varies strongly and periodically, with a period of about 20 frames, from a maximum value to a minimal one. The sequence *Marylin* is particularly complicated because of the large motion of the camera. In addition, some intensity variations of the light source are deliberately caused: around the iteration 135, the direct light is switched off, which induces some strong illumination changes.

3) Sequence *Hill*⁷ (Figure 5d) shows an outdoor scene acquired at different times of the day.

In sequences *Cylinder2*, *Planar object* and *Marylin*, $Z_P \simeq S_z$, which is

⁷This sequence can be found in the image database CMU/VASC : <http://vasc.rhcmu.edu/idb/html/motion/index.html>

not a favorable case for any photometric models (see section 4 and Table 1). In the sequence *Hill* $S_z \gg Z_P$, which is a favorable configuration. Each sequence is played forward and backward, in order to check the good behavior of the tracking, *i.e* the symmetry of residuals curve.

Results. Table 3 reports the percentage of correct points w.r.t \mathcal{N} , for each tracking technique and each sequence. P_3 tracks the largest number of points for $\mathcal{N} \leq 15$ (up to $\mathcal{N} = 25$ for *Cylinder2*). P_6 does not converge⁸ for small \mathcal{N} but performs well for $\mathcal{N} \geq 25$.

As noticed in section 6.1, the behavior of the mean residuals has also to be considered. As can be seen on Fig.6 to 9, the residuals evolve in a similar way as the illumination changes. For instance on Fig. 7, they vary periodically with the same frequency as the lighting. In most sequences, for $\mathcal{N} = 9$, P_3 provides the lowest residuals, although these values are computed from a larger number of points than J and N (see Table 3) especially in *Marylin*. P_3 offers almost similar residuals as J and N in *Planar surface* and *Hill* but it performs better when only specular highlights are considered (see Fig. 6b). In contrast, its residuals are higher when lighting changes occur (see Figures 7b and 8c) except for the *Hill* sequence (see Figure 9c), which offers a favorable configuration for the photometric models (see section 3.3.1). (see the residuals on the Fig. 7c, 8c, 9c) P_6 always yields the lowest residuals and tracks the largest number of points. As an example of the relevance of the photometric correction, Fig. 10 (a) shows the temporal evolution of the luminance of a window \mathcal{W} tracked in the sequence of Fig. 7 and Fig. 10 (b)

⁸The number of unknown parameters is too large to be properly estimated by the low number of pixels in \mathcal{W} .

shows the same \mathcal{W} after proper photometric correction by P_6 . The evolutions of the spatial photometric corrections $\boldsymbol{\alpha}^\top \mathbf{u}$ and $\boldsymbol{\beta}^\top \mathbf{u}$ in \mathcal{W} are displayed on Fig. 10 (c) and Fig. 10 (d) respectively.

Computation times. Consider a point correctly tracked by each method. The computation times are reported in Table 4, for $\mathcal{N}=9, 15$ and 35. Obviously N and P_6 are the most time-consuming techniques, either because of the computation of the photometric normalization or because of the high number of photometric parameters. In addition, they might require a larger number of iterations to converge. For $\mathcal{N}=9$ and 15, C , J and P_3 have similar computation times⁹.

6.3. Discussion

On small windows of interest, J , N and P_6 perform bad compared to C and P_3 . This can be partly explained by a sensitivity to image noise: when a pixel is noisy in \mathcal{W} , the values of μ_f , σ_f , μ_g , σ_g (see (15)), and α also become noisy since they depend on the luminance. For J and P_6 , α is multiplied by f , consequently an error on α can have a huge influence and the minimization can lead to an incorrect value of $\boldsymbol{\mu}$.

For small \mathcal{W} , P_3 tracks a larger number of points than N and J by correctly compensating for the specular highlights and lighting changes. Its performance is reduced on large \mathcal{W} under lighting changes since the model has to approximate somehow the *albedo* of the object by a first order polynomial on \mathcal{W} , which is a strong assumption when reflectance varies drastically.

Note also that, even if P_3 requires the computation of an additional pa-

⁹However, note that the KLT trackers can be accelerated on GPU units as in [32].

Table 3: Percentage of the points that have been tracked up to the end of the sequence with respect to the points which were initially selected, with regard to \mathcal{N} .

(a) *Cylinder2* (137 points selected)

\mathcal{N}	9	15	25	35
C	86.5	80.2	71.4	60.5
N	40.6	69.5	83.3	71.4
J	76.7	85.5	85.7	73.9
P_3	96.2	93.1	87.3	77.3
P_6	-	64.8	81.7	88.2

(b) *Planar object* (58 points are selected)

\mathcal{N}	9	15	25	35
C	63.8	36.2	8.6	6.9
N	77.6	91.4	96.6	93.1
J	67.2	51	87.9	89.7
P_3	100	100	96.6	96.6
P_6	48.3	94.8	100	100

(d) *Hill* (156 points are selected)

\mathcal{N}	9	15	25	35
C	49.6	20	11.2	9.6
N	45.9	63.7	73.7	75
J	55.6	70.4	85.6	93.3
P_3	74.8	75.6	86.4	95.2
P_6	-	77.8	89	97.1

(e) *Marylin* (56 points are selected)

\mathcal{N}	9	15	25	35
C	0	0	0	0
N	0	21.4	17.9	17.9
J	0	7.2	10.7	17.9
P_3	46.4	14.3	7.2	3.6
P_6	-	14.3	42.9	39.3

parameter w.r.t J and consequently the inversion of a wider matrix, the computation times of these techniques are similar, due to a better convergence of P_3 .

P_6 is the most accurate tracker for large \mathcal{W} , because the use of a comprehensive photometric model improves the estimation of the motion model during the sequence.

Table 4: Computation times (in ms) of the tracking of one point in the sequence *Cylinder2*, and planar object for $\mathcal{N}=9, 15$ et 35 . (- = no convergence)

Specular highlights				Lighting changes			
Method	$\mathcal{N}=9$	$\mathcal{N}=15$	$\mathcal{N}=35$	Method	$\mathcal{N}=9$	$\mathcal{N}=15$	$\mathcal{N}=35$
C	1.3	2.7	21	C	1.3	2.9	11.4
N	4.6	6.8	31.2	N	4.3	3.5	14.1
J	1.7	3.1	21.7	J	1.6	3.2	11.5
P_3	1.4	3.2	12.3	P_3	2	3.5	13.8
P_6	-	8.5	25	P_6	-	5.9	18.4

7. Conclusions

Through the analysis of specular reflection models, this paper has explained the assumptions on which the most widely used photometric models are implicitly based. Then, it has proposed some local photometric models which rely on a precise analysis of the reflection, and on the assumption that illumination changes vary smoothly in local areas of the image.

The proposed photometric models can be useful in many computer vision applications where lighting is not perfectly controlled, especially in outdoor experiments or for mobile cameras. This paper incorporated the models into two KLT feature point tracking procedures.

By compensating for the spatial variations of illumination changes, the proposed methods have proved to be more robust than the existing ones. The first tracker is well adapted for small windows of interest, while the

second one is applicable for larger windows of interest. Experimental results on several images sequences have shown good convergence and accuracy for these procedures. Future work will investigate the combination of the two photometric models depending on the characteristics of the point to track and depending on the local characteristics of illumination changes that are computed during the sequence.

References

- [1] A. Comport, E. Malis, P. Rives, Accurate quadrifocal tracking for robust 3D visual odometry, in: IEEE Int. Conf. on Robotics and Automation ICRA'07, 2007, pp. 40–45.
- [2] F. Espiau, E. Malis, P. Rives, Robust features tracking for robotic applications: towards 2D 1/2 visual servoing with natural, in: IEEE Int. Conf. on Robotics and Automation, ICRA'02, Washington, USA, 2002.
- [3] D. Ziou, S. Tabbone, Edge detection techniques - An overview, Int. Journ. of Pattern Recognition and Image Analysis 8 (1998) 537–559.
- [4] M. Pressigout, E. Marchand, Real-time 3D model-based tracking: Combining edge and texture information, in: IEEE Int. Conf on Robotics and Automation, ICRA'06, Orlando, Florida, 2006.
- [5] C. Harris, M. Stephens, A combined corner and edge detector, in: 4th Alvey Vision Conference, 1988, pp. 147–151.
- [6] S. Bouchafa, B. Zavidovique, Efficient cumulative matching for image registration, Image and Vision Computing 24 (1) (2006) 70–79.

- [7] B. Lucas, T. Kanade, An iterative image registration technique, in: IJ-CAI'81, Vancouver, British Columbia, 1981, pp. 674–679.
- [8] C. Tomasi, T. Kanade, Detection and tracking of point features, Technical report CMU-CS-91-132, Carnegie Mellon University (April 1991).
- [9] K. Horn, B. G. Schunck, Determining optical flow, *Artificial Intelligence* 7 (1981) 185–203.
- [10] J. Shi, C. Tomasi, Good features to track, in: *IEEE Int. Conf. on Computer Vision and Pattern Recognition, CVPR'94*, Seattle, Washington, USA, 1994, pp. 593–600.
- [11] J. Odobez, P. Bouthemy, Robust multiresolution estimation of parametric motion models, *Int. Journ. of Visual Communication and Image Representation* 6 (4) (1995) 348–365.
- [12] J. Buenaposada, L. Baumela, Real-time tracking and estimation of plane pose, in: *IEEE Int. Conf. on Pattern Recognition, Vol. 2*, Quebec, Canada, 2002, pp. 697–700.
- [13] C. Collewet, A. Alhaj, F. Chaumette, Model-free visual servoing on complex images based on 3D reconstruction, in: *IEEE Int. Conf. on Robotics and Automation ICRA'04*, New Orleans, USA, 2004.
- [14] P. Tissainayagam, D. Suter, Assessing the performance of corner detectors for point feature tracking applications, *Image and Vision Processing* 22 (6) (2004) 663–679.

- [15] S. Benhimane, E. Malis, Homography-based 2D visual tracking and servoing, *The Int. Journ. of Robotics Research* 26 (7) (2007) 661–676.
- [16] G. D. Hager, P. N. Belhumeur, Efficient region tracking with parametric models of geometry and illumination, *IEEE Trans. on Pattern Analysis and Machine Intelligence* 20 (10) (1998) 1025–1039.
- [17] T. Tommasini, A. Fusiello, E. Trucco, V. Roberto, Improving feature tracking with robust statistics, *Pattern Analysis & Applications* 2 (1999) 312–320.
- [18] S. Negahdaripour, Revised definition of optical flow: integration of radiometric and geometric cues for dynamic scene analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20 (9) (1998) 961 – 979.
- [19] H. Hausseker, D. Fleet, Computing optical flow with physical models of brightness variation, in: *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, Vol. 2, Hilton Head, 2000, pp. 760–767.
- [20] S.-H. Lai, Computation of optical flow under non-uniform brightness variations, *Pattern Recognition Letters* 25 (8) (2004) 885 – 892.
- [21] M. Black, D. Fleet, Y. Yacoob, Robustly estimating changes in image appearance, *Computer Vision and Image Understanding* 78 (2000) 8 – 31.
- [22] A. S. Georghiades, P. Belhumeur, D. J. Kriegman, From few to many: generative models for recognition under variable pose and illumination,

- IEEE Trans. on Pattern Analysis and Machine Intelligence 23 (8) (2001) 643 – 660.
- [23] P. Gros, Color illumination models for image matching and indexing, in: IEEE Int. Conf. on Pattern Recognition ICPR'00, Barcelona, Spain, 2000.
 - [24] S. Soatto, H. Jin, P. Favaro, Real-time feature tracking and outlier rejection with changes in illumination, in: IEEE Int. Conf. on Computer Vision, Vancouver, Canada, 2001, pp. 684–689.
 - [25] G. Silveira, E. Malis, Real-time visual tracking under arbitrary illumination changes, Minneapolis, USA, 2007.
 - [26] B.-T. Phong, Illumination for computer generated images, Communications of the ACM 18 (6) (1975) 311–317.
 - [27] K. Torrance, E. Sparrow, Theory for off-specular reflection from roughened surfaces, Journ. of the Optical Society of America 57(9).
 - [28] J. Lambert, Photometria sive de mensura de gratibus luminis, colorum et ombrae, Ebrhard Klett, Augsburg, 1760.
 - [29] J. F. Blinn, Models of light reflection for computer synthesized pictures, in: Proc. 4th annual Conference on Computer Graphics and Interactive Techniques SIGGRAPH '77, 1977, p. 192.
 - [30] M. Gouiffès, C. Collewet, C. Fernandez-Maloigne, A. Trémeau, Modeling illumination changes for matching correspondances, its validity

and its application to robust tracking, available on <http://hal.archives-ouvertes.fr/>, Tech. Rep. hal-00430799, IEF, Univ. Paris Sud 11, CNRS UMR 8622 (2007).

- [31] M. Gouiffès, C. Collewet, C. Fernandez-Maloigne, A. Trémeau, A photometric model for specular highlights and lighting changes. Application to feature points tracking., in: IEEE Int. Conf. on Image Processing ICIP'06, 2006.
- [32] J.-S. Kim, M. Hwangbo, T. Kanade, Realtime affine-photometric klt feature tracker on gpu in cuda framework, in: Workshop on Embedded Computer Vision (ECV), 2009.

Appendix A. Conditioning

The conditioning of the trackers detailed in that report can be compared. Indeed, each linear equation system involved in the tracking procedure (see equations (31), (33), (34), (36) and (38)) can be written as $\mathbf{A}\mathbf{x} = \mathbf{y}$ and more precisely

$$\begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{12}^\top & \mathbf{A}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} \quad (\text{A.1})$$

Consequently, the inversion of \mathbf{A} is given by

$$\mathbf{A}^{-1} = \begin{pmatrix} \mathbf{I}_3 & \mathbf{0} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{12}^\top & \mathbf{I}_3 \end{pmatrix} \begin{pmatrix} \nabla^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{I}_3 & -\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{0} & \mathbf{I}_3 \end{pmatrix} \quad (\text{A.2})$$

where ∇ is the Schur complement $\nabla = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{12}^\top$. The inversion of ∇ can be achieved in the following way:

$$\nabla^{-1} = \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}(\mathbf{A}_{22} - \mathbf{A}_{12}^\top\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}\mathbf{A}_{12}^\top\mathbf{A}_{11}^{-1} \quad (\text{A.3})$$

Since $(\mathbf{A}_{12}\mathbf{A}_{22}^{-1})^\top = \mathbf{A}_{22}^{-1\top}\mathbf{A}_{12}^\top$ we can introduce :

$$\mathbf{M} = \begin{pmatrix} \mathbf{I}_3 & -\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{0} & \mathbf{I}_3 \end{pmatrix}$$

so that

$$\mathbf{A}^{-1} = \mathbf{M}^\top \begin{pmatrix} \nabla^{-1} & 0 \\ 0 & \mathbf{A}_{22}^{-1} \end{pmatrix} \mathbf{M}$$

Consequently, the inversion of \mathbf{A} succeeds if ∇ and \mathbf{A}_{22} are well-conditioned and can be correctly inverted. ∇ is well conditioned when \mathbf{A}_{11} is well-conditioned (see (A.3)).

In the tracking techniques, for each approach the matrix \mathbf{A}_{11} is the same. Therefore, the comparison of the conditioning of the method only depends on the conditioning of \mathbf{A}_{22} . The matrices associated to the methods which approximate the photometric parameters \mathbf{A}_{22}^J (for the Jin's technique), $\mathbf{A}_{22}^{P_3}$ and $\mathbf{A}_{22}^{P_6}$ are written as:

$$\begin{aligned} J \quad \mathbf{A}_{22}^J &= \sum_m (f(m), 1)(f(m), 1)^\top \\ P_3 : \mathbf{A}_{22}^{P_3} &= \sum_m \mathbf{u}\mathbf{u}^\top \\ P_6 : \mathbf{A}_{22}^{P_6} &= \sum_m (\mathbf{u}f(m), \mathbf{u})(\mathbf{u}f(m), \mathbf{u})^\top \end{aligned} \tag{A.4}$$

The matrix $\mathbf{A}_{22}^{P_3}$ is the best well-conditioned. On the contrary, the matrix \mathbf{A}_{22}^J and $\mathbf{A}_{22}^{P_6}$ are ill-conditioned and their terms depend on the image through $f(m)$.

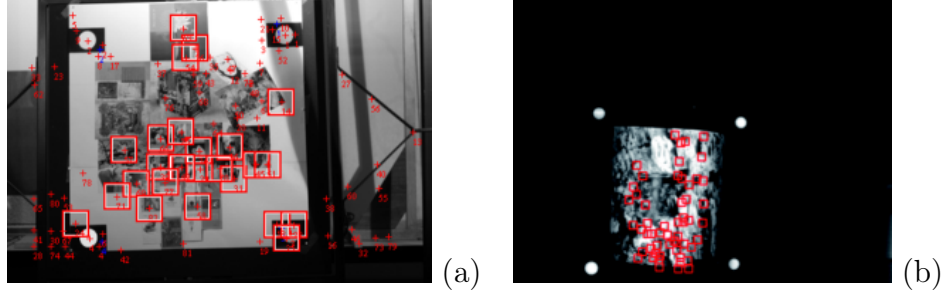


Figure 2: Sequences used to evaluate the accuracy of the tracking procedures. (a) PLANAR SURFACE. (b) CYLINDER.

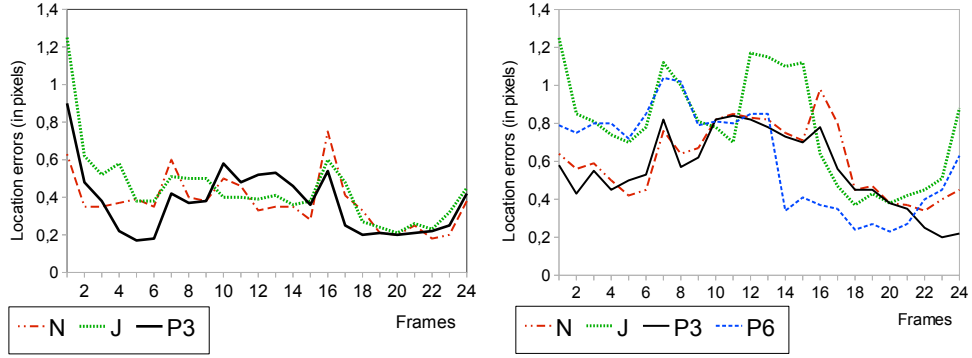


Figure 3: PLANAR SURFACE. Convergence residuals and location errors for $\mathcal{N} = 9$ (left) and $\mathcal{N} = 25$ (right).

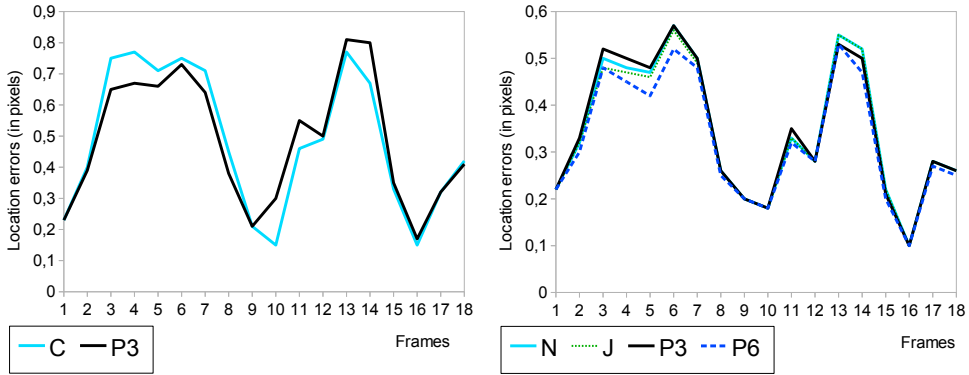


Figure 4: CYLINDER. Convergence residuals and location errors for $\mathcal{N} = 9$ (left) and $\mathcal{N} = 25$ (right).

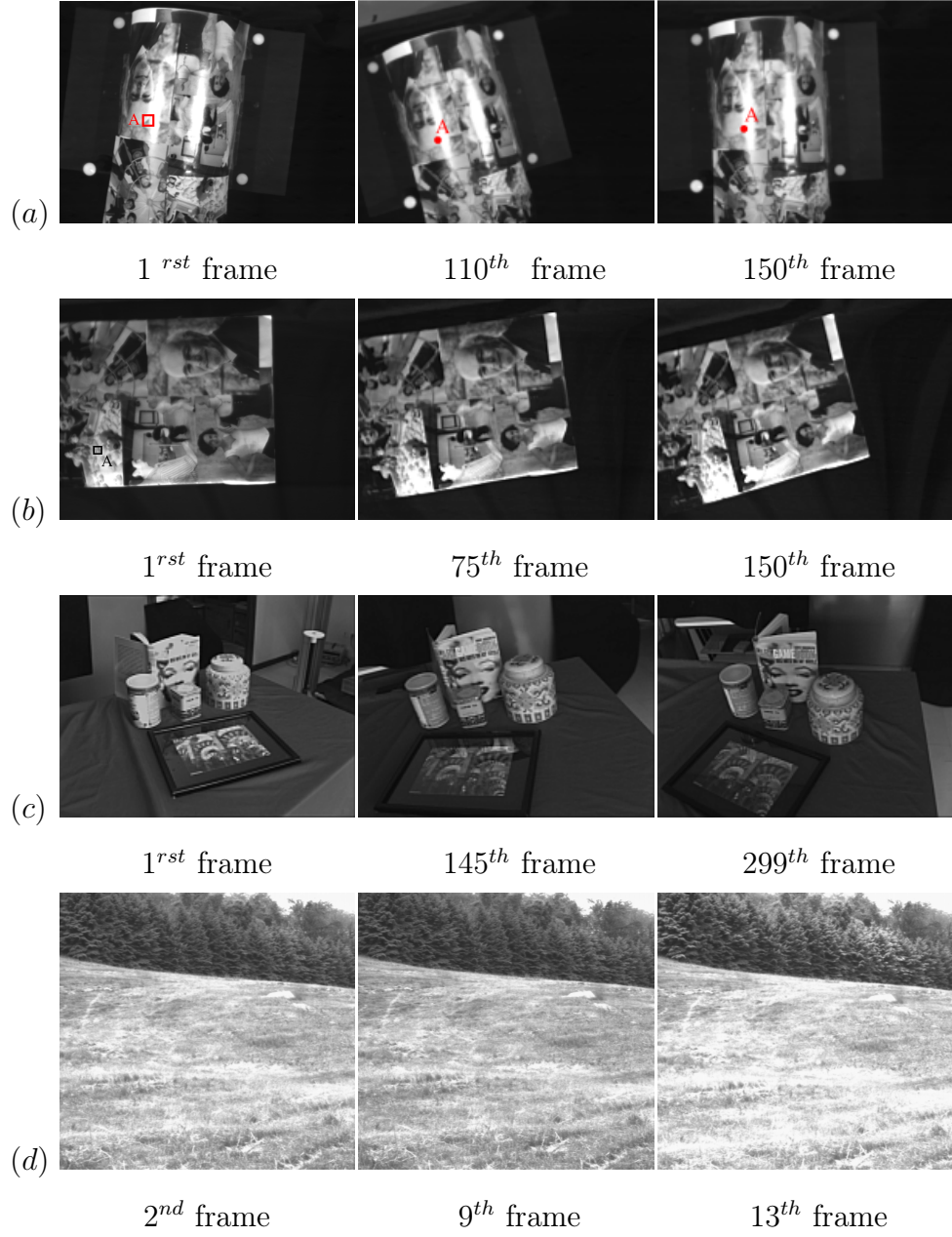


Figure 5: Images sequences. (a) *Cylinder2*. (b) *Planar object*. (c) *Marylin*. (d) *Hill*.

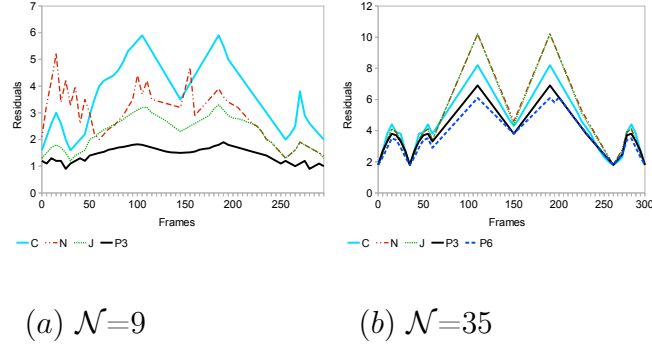


Figure 6: Sequence *Cylinder2*. 37 points have been selected. Average of the residuals with regard the frame number, with a window size (a) $\mathcal{N} = 9$ (b) $\mathcal{N} = 35$.

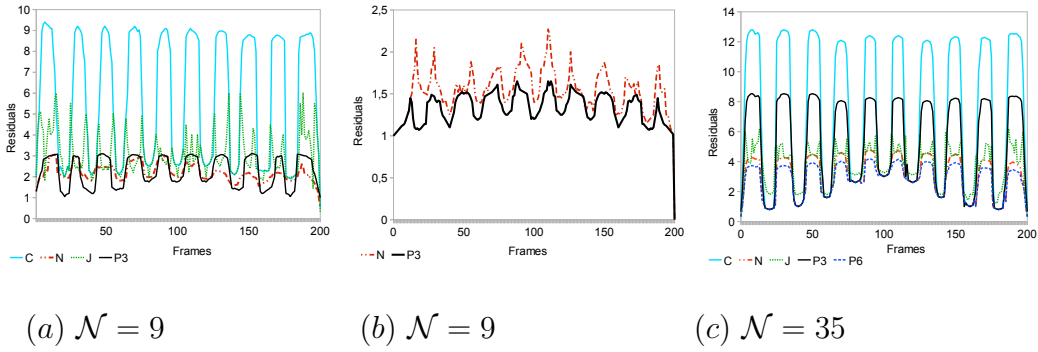


Figure 7: Sequence *Planar object*. (a) Average residuals obtained with $\mathcal{N}=9$. (b) Residuals obtained with $\mathcal{N}=9$ on the points which are tracked simultaneously by N and P_3 . (c) Average residuals obtained with $\mathcal{N}=35$.

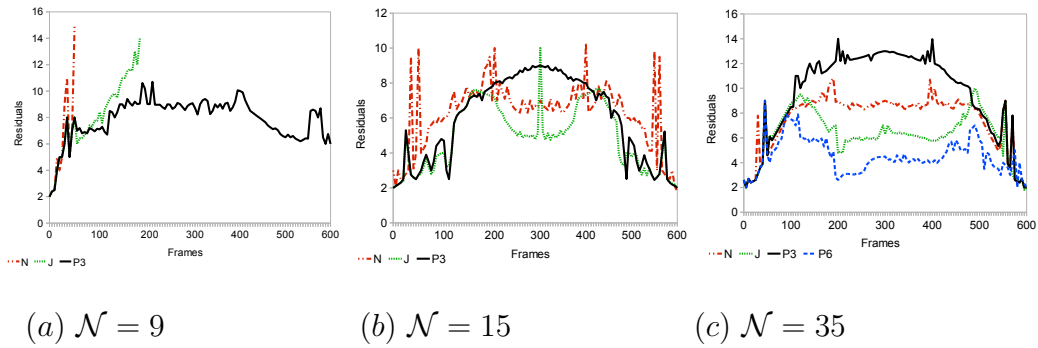


Figure 8: Sequence *Marilyn*. Average residuals obtained with (a): $\mathcal{N}=9$, (b): $\mathcal{N}=15$ and (c): $\mathcal{N}=35$.

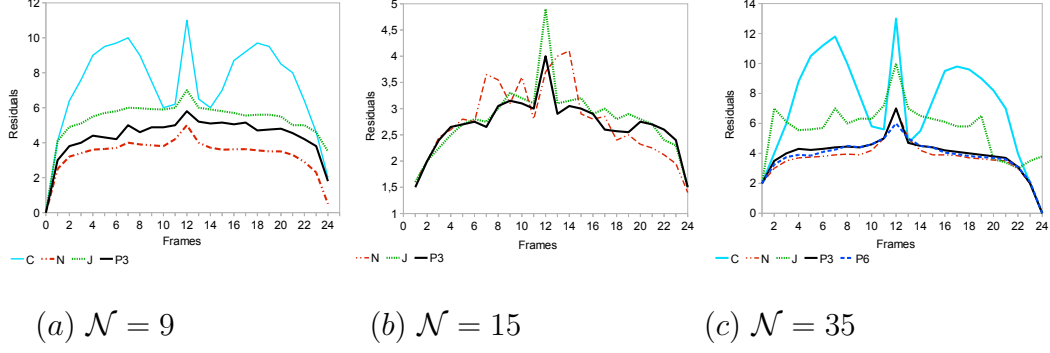


Figure 9: Sequence *Hill*. (a) Average residuals obtained with $\mathcal{N}=9$. (b) Residuals obtained with $\mathcal{N}=9$ on the points which are tracked simultaneously by N and P_3 . (c) Average residuals obtained with $\mathcal{N}=35$.

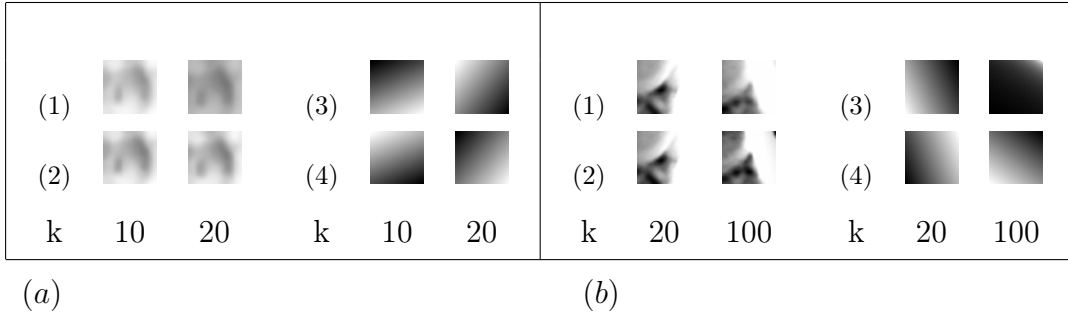


Figure 10: Examples of photometric correction in the sequences (a) *Planar Object* and (b) *Cylinder2* (point A) for different times k : (1) without photometric correction. (2) correction with P_6 . (3) Evolution of $\alpha^\top \mathbf{u}$. (4) Evolution of $\beta^\top \mathbf{u}$.